Mapping Web Space with the Issuecrawler

Richard Rogers

Abstract

Changes in the conceptualization of Web space may be gleaned from the kinds of visualizations made over the past decade. The piece concerns itself with the visualization modules made for one Web 'mapping' device, the Issuecrawler network location software. It briefly periodizes understandings of Web space by examining the contexts in which Issuecrawler mapping modules were conceived and built: the site inlink list (in the Web as hyperspace period), the circle map or virtual roundtable (in the period of the Web as neo-pluralistic space), the cluster or issue network map (in the Web as network period) as well as the geographical map or the distributed geography of an issue (in the current locative period).

The piece concludes with a series of concrete research projects. Together with allied scraping tools, the Issuecrawler has been employed, among other purposes, for censored Website discovery, the analysis of the policy impact of NGOs, and an understanding of the interests served by holding an event. With the studies, the focus moves from the 'metaphysics' of software-made space (sphere, network) to the specific info-political geographies that can be charted with the aid of the tools. The Web becomes a space to view Iranian censorship policy, media justice actor and issue recognition by a U.S. governmental agency, and Western information policy circles' interest in Dutch hacker agendas.

Introduction: The Death of Cyberspace

The symbolic end of 'cyberspace' may be located in the Yahoo lawsuit in May 2000, brought before the Tribunal de grande instance de Paris by two French non-governmental organizations, the French Union of Jewish Students and the League Against Racism and Anti-Semitism. The suit ultimately led to the ruling in November 2000 that called for software to block Yahoo's Nazi paraphernalia pages from Web users located in France.

Web software now routinely knows a user's geographical location, and acts upon the knowledge. You are reminded of the geographical awareness of the Web when you type into the browser, google.com, and are redirected to google.fr. Whilst it may be viewed as a practical and commercial effort to connect users with languages and local advertisements, the search engine's IP-geo-location handling also may be viewed as the demise of cyberspace as place-less space [MIL01].[1] With location-aware Web devices, cyberspace becomes less an experience in displacement than one of re-placement – you are sent home by default.

The announcement of the death of cyberspace through the revenge of geography has consequences for any theorising of the history of Web space. The question posed here concerns how Web space is conceptualized by devices that have sought to 'map' the Web, especially without employing conventional politico-geographical cartography or borrowing from geological metaphors, such as thematic islands, peaks or valleys [DOD01]. In the following I treat one device in particular. The Issuecrawler's sense of Web space is explored through a brief history of the visualization modules created for the software – a history that also seeks to periodize understandings of Web space. It does so through a reflection on how the visualizations provided commentary on contemporaneous Web thought.

The Issuecrawler is server-side Web network location software. Input URLs into the Issuecrawler, and the software crawls the URLs, captures page/site outlinks, performs co-link analysis, and outputs the results in lists as well as visualizations. The software was conceived in the mid-1990s at the Department of Science and Technology Dynamics, University of Amsterdam [ROG96], and has a forerunner in the Netlocator, also known as the De-pluralising Engine, built in Maastricht during the Jan van Eyck Design and Media Research Fellowship, 1999-2000.

[1] Web ethnographers pointed out as long ago as 2000 that geography matters more than the medium. The local and the national, it has been argued, exert more pull than the special medium-specific practices of the alleged placeless space.

Tethering Websites in Hyperspace (by Inlinks)

The Netlocator began with the insight that Websites (or Webmasters) link selectively as opposed to capriciously. There is a certain optionality in link-making. Making a link to another site, not making a link, or removing a link, may be viewed as acts of association, non-association or disassociation, respectively. Later, we learned through a Georgia Tech study and our own observations and interviews, that hyperlinks are matters of organizational policy, especially for corporations and government [KEH99], [GOV99].

Selective link-making could create space when one conceives of space as that demarcated by limited acts of association. The demarcationist approach performs an important break with cyberspace by suggesting that hyperlinking behaviors dismantle the 'open-ended-ness' of cyberspace, one that informed the idea of 'placeless-ness.'

What types of associations are on display in hyperlinks ('reading between the links'), and what could be the shapes of spaces demarcated by link associations? In the late 1990s and early 2000s the leading visualizations we discussed were the Plumb Design's ThinkMap Visual Thesaurus as well as the I/O/D's WebStalker, followed shortly thereafter by TouchGraph's Google Browser as well as Theyrule.net by Josh On. All are non-directed graphs, without arrowheads, which is to say that the elements (synonyms, site pages, board members and companies) are associated (and lines are drawn between them), without specifying a uni- or bi-directional association. Undirected graphs, arguably, derive from a path model of the Web, also built into browsers (with the forward and backward arrows), and lead to ideas about every link being a two-way link [NEL99], [BER99]. Seeing the Web in terms of paths is not far-fetched, since one may surf from page to page, and use the browser buttons, or the browser history, to retrace one's steps and also move 'forward' again. However, on the Web, two-way links, it may be observed, are less frequent than one-way links. Viewing any hyperlink as a bi-directional association, we learned at the time, also has its infamous cases, whereby for example a German ministerial site was accused of being linked to a call boy network [MAR00]. The Bundesministerium fuer Frauen und Jugend linked to a women's issues info site, and that info site linked to a call boy network. To the *Bild Zeitung* this Web path implicated government.

To stand on the shoulders of Vannevar Bush, Theodore Nelson and other path and hypertext model pioneers would view the Web as pathway space (for the surfer) [BUS45], [MAR99]. The Netlocator (and later the Issuecrawler), however, strove to distance itself from the Web as pathway space, and instead concentrated on the Web as selective associational space (made by Webmaster linking). How does one view associations? As is well-known, a site's outlinks, most readily in the form of one or more link or resources lists, are viewable to a site visitor. To gain a sense of a site's inlinks, however, requires the use of the advanced search of an engine, or access to the referrer logs of a site. Until the creation of 'trackback,' a feature implemented in the Movable Type blogging software in 2002 that shows backlinks to a posting, inlinks in the late 1990s were not an everyday concern. Only ranking algorithm makers, most notably Google with the PageRank system, made use of them. Nowadays, on the Web as well as in the blogosphere and in online news, devices recommend pages routinely by counting inlinks, e.g., 'most blogged' stories at the *New York Times* and the *Washington Post*. In all, concern with inlinks as a marker of page relevance or reputation marked a major shift in the underpinnings of Web space.

Counting inlinks addressed the site authority problem. Previously, in the mid-nineties the foremost issue concerning search engine developers related to how to separate the 'real name' from the borrowers of the name, e.g., to return Harvard University at the top of the list when Harvard is queried. In leading search engine results (AltaVista's), the "eminent scientist and the isolated crackpot [stood] side by side," as one leading author put it [RHE94]. In their ranking logics, AltaVista granted site owners the authority to describe the content of their sites (in metatags) and their descriptions became the basis for the engine returns. The Web became a space displaying "side-by-side-ness," fitting with contemporaneous ideas about its pluralizing potential [BAR96]. Google, conversely, granted other sites that authority (hyperlinks and link pointer text). Counting inlinks and having other sites grant authority through linking (and naming their links well) form the basis for most search engine algorithms these days, including Yahoo's as well as MSN's. Once a major competitor to automated search engines, the directory has declined.

De-pluralizing the Web

Which links should be counted? The observation we made in the late 1990s was that search engines' 'population' for link-counting was the entire Web (or the percentage of it they were able to index). Instead of focusing on what the 'most influential' [social network metric] are calling the carriers of the term, on the record on the Web (which is how I would summarize the dominant search engine ranking algorithms), we preferred seeking what could be called 'organizational networks.' Insert a set of URLs of organizations working in the same area, and return those organizations (or URLs) which have received at least two links from the starting points.

Thus, like Google's for the entire Web, the Netlocator's (and, later the Issuecrawler's) algorithm for a portion of the Web crucially sought to take into account sites' inlinks. Once the crawling and co-link analytical procedure of the Netlocator completed, a list of sites in the network (the results) were displayed, color-coded as governmental (.gov), commericial (.com), non-governmental (.org), and scientific (.edu), including country-specific sub-level domains. (.Ac.uk, for example, would count as scientific.) When an actor was clicked, the links it *received* were highlighted. It was not just inlink counts, but types of inlinks, that concerned us, however. When showing an actor's inlink types, the Web could be made into an actor reputational space by showing which links a site received.

How and why do sites link [PAR05]? The Netlocator-related link language for outlinks and inlinks provided a schematic for linking behavior generally, according to domains as well as further qualitative taggings. (See Figure One.) In one of our first extended case studies, on genetically modified food, inlinks and outlinks provided actor profiles according to types of links received and given. For example, three corporate sites were compared; the sites' respective standings differ according to the types of links received, and sites' respective display of awareness according to types of links given. One corporation has a different standing by virtue of receiving links from NGOs and government, as opposed to from other corporations only (see Figure Two). More qualitatively, links were classified as cordial, critical or aspirational. Cordial links are the most common – to project partners and affiliates and other friendly or respected information sources. Critical links, largely an NGO undertaking, have faded in practice, and aspirational links are made normally by smaller organizations to establishment actors, often by organizations desiring funding or affiliation (see Figure Three).

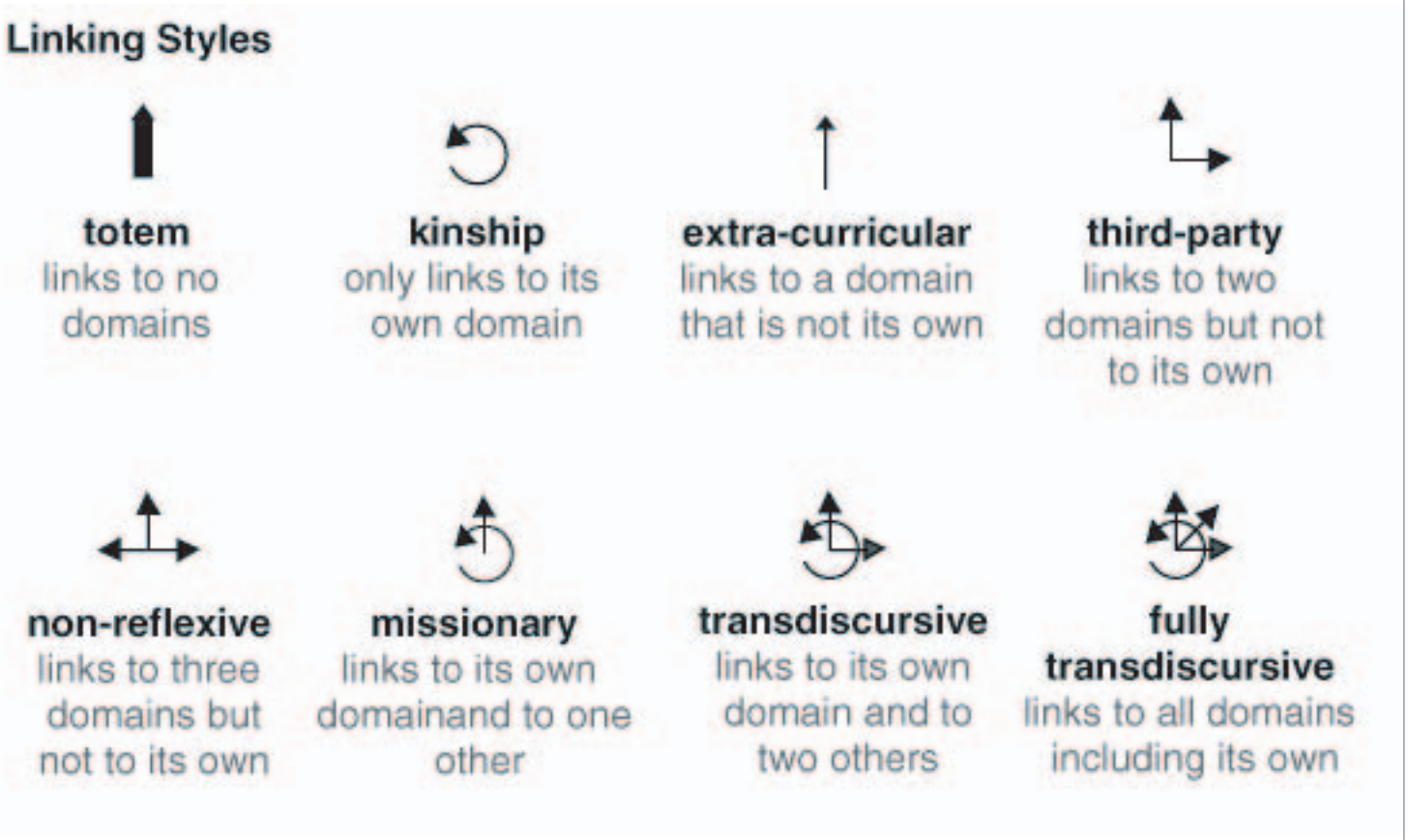


Figure One: Actor Hyperlink Language, Govcom.org, Design and Media Research Fellowship, Jan van Eyck Academy, Maastricht, 1999.

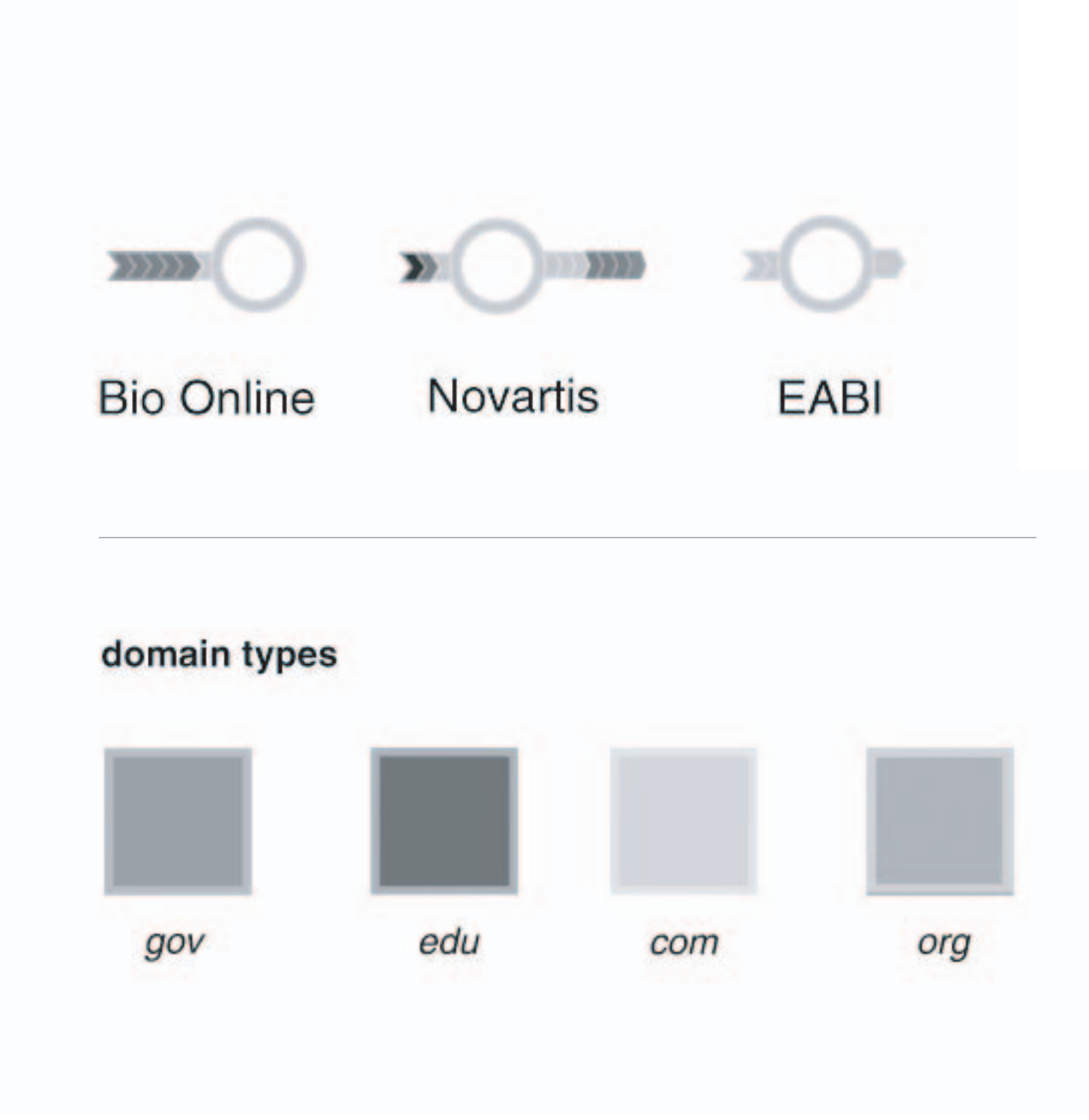


Figure Two: Actor Reputational Profiles by Inlink and Outlink Types. Govcom.org, Design and Media Research Fellowship, Jan van Eyck Academy, Maastricht, 1999.

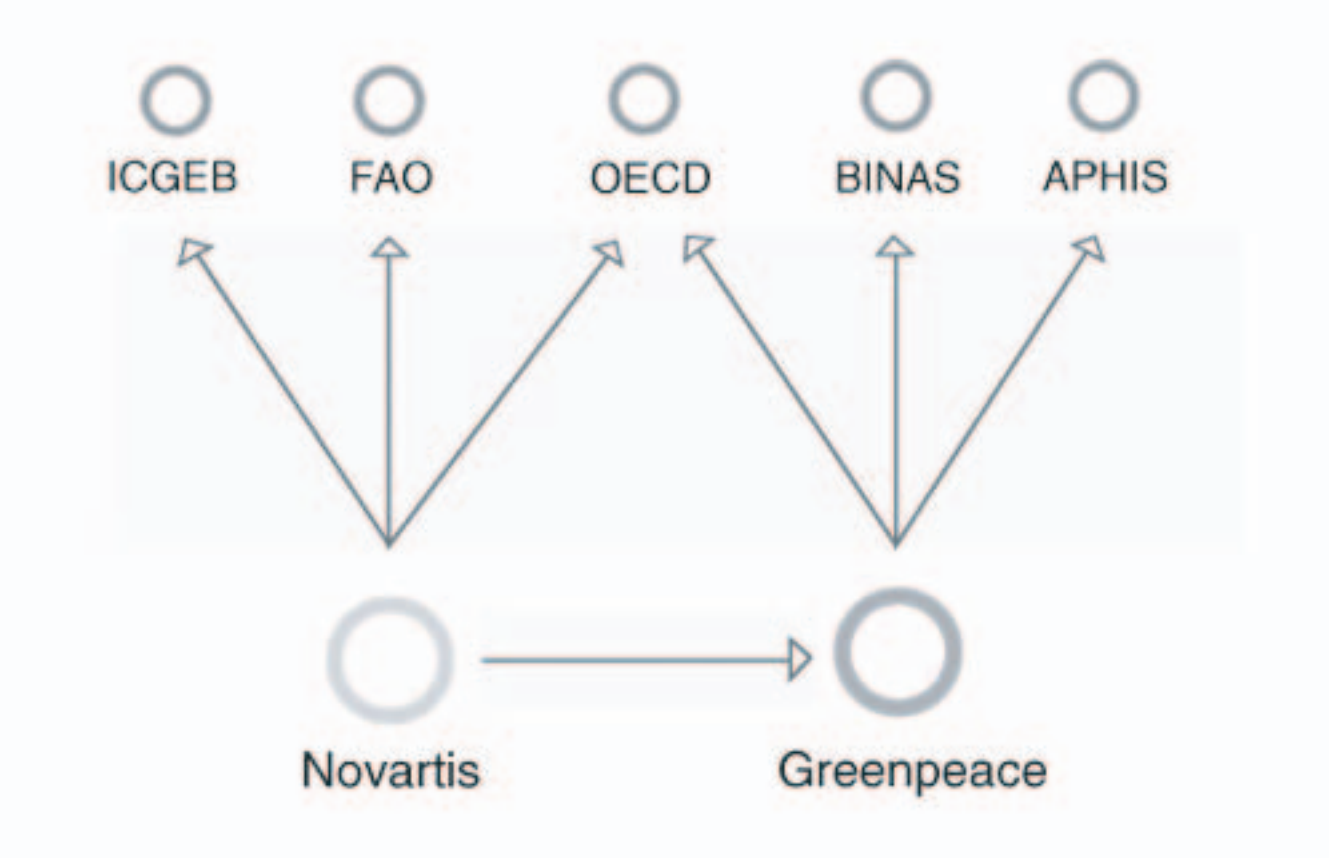


Figure Three: Aspirational linking in the GM Food Issue Space. Novartis links to Greenpeace. Greenpeace does not link back. Greenpeace and Novartis link to government. Government does not link back. Govcom.org, Design and Media Research Fellowship, Jan van Eyck Academy, Maastricht, 1999.

In pre-Netlocator discussions in 1998, the shape given to the visualization of associational linking in Web space was based, initially, on astronomical charts. Generally, thinking in terms of the Web as a universe (to be charted) coincided with early ideas of the Web as a hyperspace, where one would jump from one site to another at a great distance. Google's "I'm feeling lucky" button also played upon the trope of hyperspace and the famed hyperspace button (from the Asteroids arcade game by Atari, released in 1979). In the period of starry night site backdrops and random site generators, Web sites, arguably, appeared untethered, individual stars whose relationships could be charted (and constellations or configurations perhaps named).

With Netlocator output (lists with interlinkings between pages/sites), the circle maps we made also evoked the 'sphere' of public sphere theory. The idea that the Web was or could be made into a pluralizing space, where familiar hierarchies of credibility may be challenged, became the focus of our visualization work. With the GM Food (1999) and Russian HIV-AIDS maps (2000), we sought to show interlinkings between sites in a kind of virtual roundtable [ROG00]. What if the Web, according to network inlink counts, were to determine who would sit at the table, instead of more familiar agenda-setters? Significantly, however, not all the actors had the same standing at the table -- some receive more links than others and thus grow larger in size. The links between the actors are considered to be entanglements. Are the linked actors all on the same side? Would only the largest nodes speak, and the smaller ones keep still? Thus our roundtable was not flat; it had complications, which we sought to capture in the notion of the 'De-pluralizing Engine,' the other name for the Netlocator. In fact, the Web should not be seen as a pluralizing space by itself, for it is creating hierarchies through inlink counts generally, and through inlink counts from most influential actors (the basis of Google's PageRank).

In the circle maps, especially those auto-generated in what came to be known as the vanilla version of the Issuecrawler (2001), the de-pluralizing spirit continued with built-in notions of a core network and a periphery, where the latter, called the 'waiting room,' comprises those actors (or sites/pages) not quite receiving enough links to sit at the table. A variation on aspirational linking was in evidence in the visualization, as it showed only links from periphery to core, and not from core to periphery. Thus the peripheral link showed a desired belonging to the core, as of yet unachieved owing to lack of sufficient inlinks.

Referring to the mapping practice, Noortje Marres prefaces her PhD dissertation with the following remark: "When we [took] to the Web to study public debates on controversial science and technology, we [found] issue networks instead" [MAR05]. Notions of the Web as debate space, with the virtual roundtable construct (however much we strove to complicate it), did not fit with the empirical findings. Even when we endeavored to *make* the Web into a debate space, by harvesting text from organizations' specific, issue-related deep pages, we found only statement juxtapositions – comments by organizations on a particular statement, but scant inter-organizational exchange (see Figure Four). Organizations would release views on an issue on their Websites (which we would capture), but forums and other dialogue spaces were not used by what could be construed as the parties to a debate. The Web could not stand in for a building – or an event where debating parties could gather. (Certain authors also began to discuss our work as evidence that the Web (or Net) should not be construed as a public sphere [DEA02].) The Web as neo-pluralistic space had come to an end.

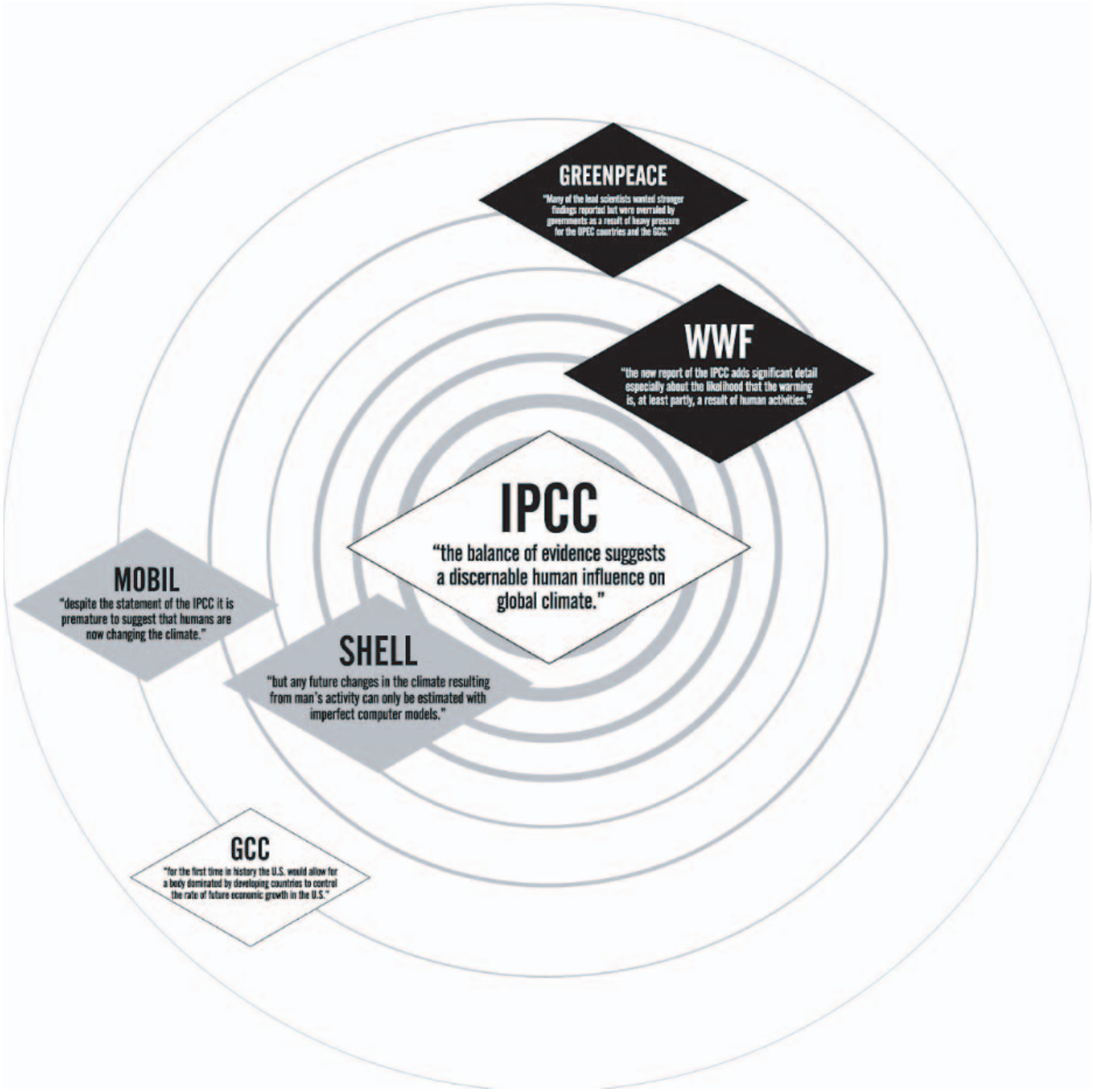
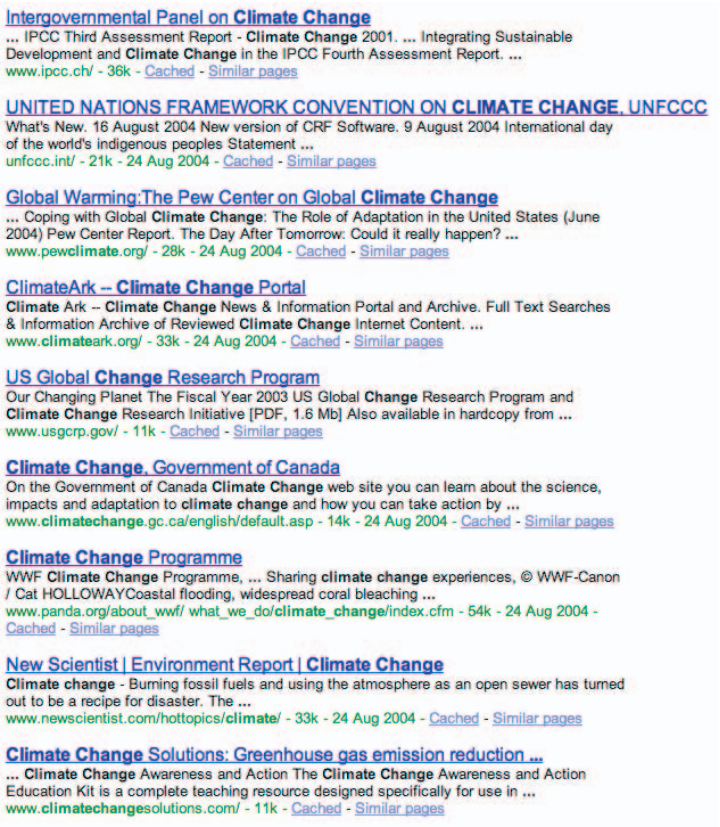


Figure Four: Key Statement in Context Map. Discursive affinities (or non-affinities) between organisations in the use of the Intergovernmental Panel on Climate Change's finding: "The balance of evidence suggests a discernible human influence on global climate", Noortje Marres, Richard Rogers and Noel Douglas, 1998.

Not only with the circle maps but also with the cluster mapping module made by Andrei Mogoutov and aguidel.com, the fully designed version of the Issuecrawler (2003), with instructions of use, described itself as an 'issue network' location and visualization machine. That the Web would come to be thought of in terms of a network space, as opposed, for example, to a virtual or online community space, relies initially on the difference between multiple- and single-site analysis. ('Online communities' these days have the tendency to be geographically concentrated and located on a single site, as Hyves in the Netherlands, Facebook in the United States, Orkut for Brazil, Cyworld in South Korea and Lunarstorm in Sweden.) When performing mulitiple site analysis, with the Issuecrawler, the crawling and co-link analysis return the sociable and the under-socialized, so to speak, in the same space. (Thus we achieved a new form of 'side-by-side-ness.') In terms of types of associations (found in Web space and network mapping more specifically), issue networks may be distinguished from popular understandings of networks, and social networking, in that the individuals or organizations in the network neither need be on the same side of an issue, nor be acquainted with each other (or desire acquaintance) [MAR06]. Actors may be antagonistic, oppositional, adversarial, unfriendly, estranged. Additionally, unlike social networks, issue networks do not privilege individuals and groups, as the networks also may be made up of a news story, a document, a leak, a database, an image or other such items, found on individual pages of Websites. (Thus the Issuecrawler considers 'deep pages' as significant for the study of issue networks. I return to this in the media justice case study below.) Taken together these actors and 'argument objects' serve as a means to interrogate the state of an issue either in snapshots or over time. 'Issue states' may be gauged, initially, by taking note of the network's actor composition [ROG04]. To take an elementary example, compare queries made in Google in 2004 for "climate change" and for "RFID". Note the actors represented in the top returns. For "climate change" we note UN scientists, governmental agencies, and other establishment actors. In the RFID (radio frequency identification) returns, we have trade press, corporations, lone activists and electronics tinkerers. By comparing the networks' actor composition (from the inlink-count-based engine returns), climate change presents itself as a far more mature issue than RFID (See Figure Five).

It is important to emphasize the Web's capacity (with the Issuecrawler) to display configured, professional and publicized culture. The networks or lists that are located rely on public displays of connection (hyperlinks), rather than informal, quiet or old-boy relationships [HOB03]. Indeed, network mapping often has as its goal to make things visible, to reveal non-public relationships, even to dig for dirt. (A 2002 search engine query resulted in the newspaper headline: "UN weapons inspector is leader of S&M sex ring" [DAI02].) Understandings of the Web as network space, together with the return of the informality of the Web (particular through the blogosphere), have given rise to an investigative outlook. The impulse relates to the Web's street proximity, its closeness to the ground, including the "fact-checking," evidential spirit of the political blogosphere.



Intergovernmental Panel on **Climate Change**
... IPCC Third Assessment Report - **Climate Change** 2001. ... Integrating Sustainable Development and **Climate Change** in the IPCC Fourth Assessment Report. ...
www.ipcc.ch/ - 36k - Cached - Similar pages

UNITED NATIONS FRAMEWORK CONVENTION ON **CLIMATE CHANGE**, UNFCCC
What's New. 16 August 2004 New version of CRF Software. 9 August 2004 International day of the world's indigenous peoples Statement ...
unfccc.int/ - 21k - 24 Aug 2004 - Cached - Similar pages

Global Warming:The Pew Center on Global **Climate Change**
... Coping with Global **Climate Change**: The Role of Adaptation in the United States (June 2004) Pew Center Report. The Day After Tomorrow: Could it really happen? ...
www.pew**climate**.org/ - 28k - 24 Aug 2004 - Cached - Similar pages

ClimateArk -- **Climate Change** Portal
Climate Ark -- **Climate Change** News & Information Portal and Archive. Full Text Searches & Information Archive of Reviewed **Climate Change** Internet Content. ...
www.**climate**ark.org/ - 33k - 24 Aug 2004 - Cached - Similar pages

US Global **Change** Research Program
Our Changing Planet The Fiscal Year 2003 US Global **Change** Research Program and **Climate Change** Research Initiative [PDF, 1.6 Mb] Also available in hardcopy from ...
www.usgcrp.gov/ - 11k - Cached - Similar pages

Climate Change, Government of Canada
On the Government of Canada **Climate Change** web site you can learn about the science, impacts and adaptation to **climate change** and how you can take action by ...
www.**climatechange**.gc.ca/english/default.asp - 14k - 24 Aug 2004 - Cached - Similar pages

Climate Change Programme
WWF **Climate Change** Programme, ... Sharing **climate change** experiences, © WWF-Canon / Cat HOLLOWAYCoastal flooding, widespread coral bleaching ...
www.panda.org/about_wwf/ what_we_do/**climate_change**/index.cfm - 54k - 24 Aug 2004 - Cached - Similar pages

New Scientist | Environment Report | **Climate Change**
Climate change - Burning fossil fuels and using the atmosphere as an open sewer has turned out to be a recipe for disaster. The ...
www.newscientist.com/hottopics/**climate**/ - 33k - 24 Aug 2004 - Cached - Similar pages

Climate Change Solutions: Greenhouse gas emission reduction ...
... **Climate Change** Awareness and Action The **Climate Change** Awareness and Action Education Kit is a complete teaching resource designed specifically for use in ...
www.**climatechange**solutions.com/ - 11k - Cached - Similar pages

RFID Journal
The only source for timely, objective news and information about **RFID** and its many business applications. ... Putting **RFID** Know-How to Work. ...
www.**rfid**journal.com/ - 26k - 23 Aug 2004 - Cached - Similar pages

RFID Journal - Frequently Asked Questions
For answers to frquently questions about **RFID** technology and consumer privacy, see Privacy FAQs. ... What is **RFID**? How does an **RFID** system work? ...
www.**rfid**journal.com/article/articleview/207 - 95k - 23 Aug 2004 - Cached - Similar pages
[More results from www.rfidjournal.com]

AIM - The global trade assocation for automatic identification
... Covering technologies such as barcode, radio frequency Identification (**RFID**), card technologies (magnetic stripe, smart card, contactless card, optical card ...
www.aimglobal.org/technologies/**rfid**/ - 37k - 23 Aug 2004 - Cached - Similar pages

Spychips: **RFID** Privacy Website
Order Katherine Albrecht's video "**RFID**: Tracking Everything Everywhere." Sign up for our free weekly newsletter ... **RFID** PRIVACY ISSUES AND NEWS. What is **RFID**? ...
www.spychips.com/ - 17k - Cached - Similar pages

RFID, Inc. Radio Frequency Identification Products....!
RFID Inc. features radio frequency identification products, also offering ISO standard interchangeable products. ... Looking for a 5 or 10 cent **RFID** Tag? ...
www.**rfid**inc.com/ - 15k - Cached - Similar pages

RFID tags: Big Brother in small packages | Perspectives | CNET ...
RFID tags: Big Brother in small packages | CNET News.com's Washington watcher Declan McCullagh says that retailers may love the concept of miniscule ...
news.com.com/2010-1069-980325.html - 46k - Cached - Similar pages

RFID - Wikipedia, the free encyclopedia
RFID. From Wikipedia, the free encyclopedia. Radio frequency identification (**RFID** ... [edit]. History of **RFID** tags. Perhaps the first work ...
en.wikipedia.org/wiki/RFID - 35k - Cached - Similar pages

EPIC **RFID** Privacy Page
EPIC logo. Radio Frequency Identification (**RFID**) Systems. ... 1). Introduction. Radio Frequency Identification (**RFID**) is a type of automatic identification system. ...
www.epic.org/privacy/**rfid**/ - 37k - Cached - Similar pages

RFID Tags in New US Notes Explode When You Try to Microwave Them
RFID Tags in New US Notes Explode When You Try to Microwave Them Adapted from a letter sent to Henry Makow Ph.D. Want to share an event with you, that we ...
www.prisonplanet.com/022904**rfid**tagsexplode.html - 40k - Cached - Similar pages

The **RFID** Weblog - **rfid**.weblogsinc.com
... A New Alternative to Bluetooth/**RFID**/NFC on the Horizon. ... FDA Looking at Wireless/**RFID** Solutons to Combat Against Counterfeit Drugs. ...

Figure Five: Issue Maturity Indication by Actor Composition Returned by Search Engine. Google Query Results for Climate Change and RFID, 24 and 23 August 2004, respectively.

The Issuecrawler takes into account a sense of a public 'real' – evinced in the making and displaying of a hyperlink. Thus, importantly, the Issuecrawler does not map what is commonly understood as 'virtual space' (as an online game environment). A map of a virtual space would be to a computer or video game what a 'site map' is to a Website, showing the world (or the pages) that have been built and how one may navigate them.

Since 2005 the Issuecrawler has been considered a mapping device for issue professionals and researchers working with that made public. In an effort to make the Issuecrawler's sense of the 'real' even less virtual, the latest visualization module, the Issuegeographer, strives to ground networks (see Figure Six). We placed issue network actors on a geographical map to show the proximity (or lack thereof) of the places of actors to the places of issues.

The focus of the visualization work began to consider actor mobility, whether networked actors move from issue to issue (or whether issues move from network to network). The provocative question read: Do networks form around issues, or are there networks in place that assume issues as they arise? Previously, in social movement research, the idea was mooted that there is 'free-floating movement potential,' in the sense of a given collection of publics which are able to form a movement, with particular conditions. That is, movements are not spontaneous uprisings, but rather more structural phenomena. May the same be said of networks? Are networks simply there, like Websites under construction, waiting for content?

Especially global issues may have typical discursive homes, as at (recurring) conferences, summits and other gatherings. Thus, we asked, is there a difference between where is an issue is happening, and where it is currently based? The notion of the 'base of an issue' takes as its point of departure professional circulation, which results in people asking each other, not where you are from, but where you are currently based. With the Issuecrawler in tandem with the Issuegeographer, the Web becomes a space where one can locate where an issue is based [GOV05].

At the time of writing the Issuecrawler is approaching 1,000 registered users, whose projects vary considerably. In the following, I present a series of concrete Issuecrawler case studies performed in 2005 and 2006: censored Website discovery (Iran), the policy impact of media justice organizations in the U.S.A., and the relationship between hackers (and hacker events) and information policy circles, answering the question as to why hold an event. In each case the Web becomes a space, with tools, that puts on display info-political geographies: Iranian censorship policy, actor and issue recognition by a U.S. governmental agency and Western information policy interest in a Dutch hacker event.

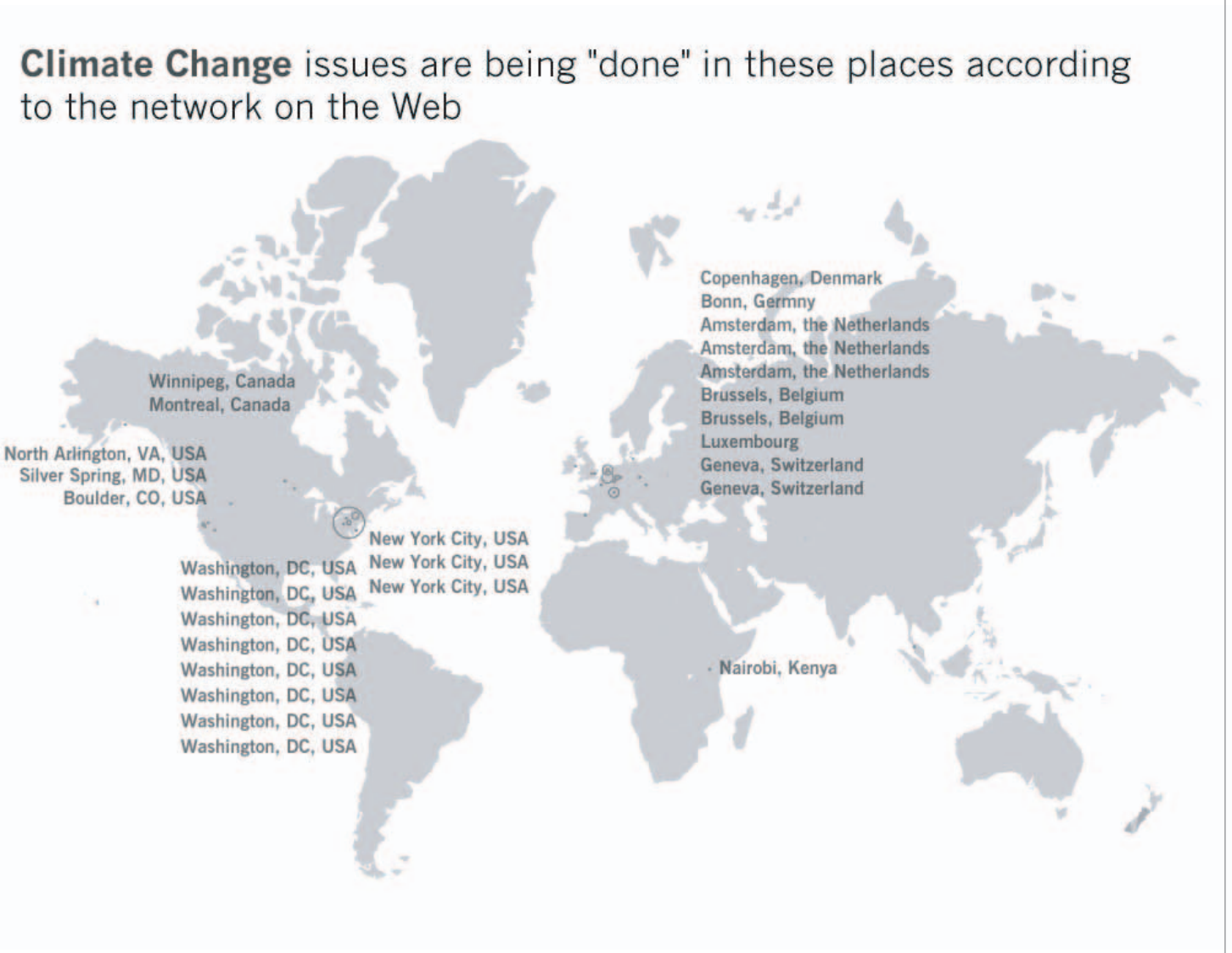


Figure Six: The Base of an Issue. Issuecrawler Results Plotted to the Issuegeographer, 2005

Case Study: Discovering URLs with the Issuecrawler:
Internet Censorship Case Study

The OpenNet Initiative, at the University of Toronto, Harvard University and the University of Cambridge, researches Internet censorship in nearly 40 countries. Among ONI's heuristic procedures, experts located in countries that censor the Internet compile lists of 'high-impact' URLs, for example, the Free Vietnam Alliance (http://www.lmvntd.org/), sub-categorized as dissident site, and the International Movement on Religious Freedom and Human Rights for Vietnam (http://www.tudotgvn.org/), sub-categorized as a human rights site [ONI06]. The URLs are fetched in browsers within the countries in question (in homes, Internet cafes, hotels, etc.), and filtering behavior, per Internet Service Provider (ISP), is recorded. The Free Vietnam Alliance is blocked when accessed through both FPT (Vietnamese Corporation for Financing and Promoting Technology) and VNPT (Vietnam Posts and Telecommunications Corporation), whilst the International Movement on Religious Freedom and Human Rights for Vietnam is accessible via VNPT and blocked by FPT.

A small paragraph on the Reporters without Borders site (rsf.org), where one may also read about the "enemies of the Internet," prompted a methodological contribution to Internet censorship research, with the Issuecrawler. In the *Handbook for Bloggers and Cyberdissidents*, in a section called "Internet-censor world championship," sub-section Saudi Arabia, it reads:

> The official Internet Service Unit (ISU) is proud to tell you it's barred access to nearly 400,000 sites and has even posted a form online for users to suggest new websites that could be blocked. The ISU says it filters sites to shield citizens from offensive material violating Islamic principles and social norms.[PAI05]

As of December 2005 the OpenNet Initiative had a list of 1966 sites to be checked in Saudi Arabia, short of the 400,000 claimed to be censored, if the referenced ISU figure approximates the situation. How is one able to discover further relevant URLs for testing of filtering behavior in a country, as Saudi Arabia?

In March 2006 the Issuecrawler was used for "dynamic URL sampling," an initial step in censored Website discovery, prior to the on-the-ground page-fetching (or the less preferred proxy server testing means). The ONI list of "political, social and religious sites," to be checked for censorship in Iran, was inserted in the harvester of the Issuecrawler, which strips URLs from text, providing a clean list of 128 URLs, the number of sites on the ONI category list. The Issuecrawler was set to one iteration of co-link analysis (by page), whereby each URL on the list is crawled, its external links (outlinks) are fetched, and those outlinks receiving at least two inlinks are retained. The starting points, or seed URLs, were 'privileged,' whereby they are kept in the results as long as they receive at least one inlink from the co-linkees. (The Issuecrawler does not retain isolates.) In this case, the purpose of privileging the starting points is to demonstrate the Issuecrawler method on the resulting map (or graph) that is outputted. One is able to view the links between seed URLs and co-linked sites on a spring map. (The node size depends on the quantity of links received, and nodes grow closer to each other when ties are stronger, as in more frequent.) Here, however, the size and placement of URLs on the graph are less relevant than the resulting URL list, for the project concerns URL discovery. Which sites do "high-impact" political, social and religious sites link to (at least twice)? (The co-link analysis becomes a site relevance threshold.)

The Issuecrawler outputted 104 nodes, which on the interactive scalable vector graphic (SVG) map shows truncated URLs, as bbc.co.uk, and on click opens the page on the site receiving the most inlinks, in this case: http://www.bbc.co.uk/persian/. One of the important discoveries made concerned the Persian-language page of the BBC World Service. Whilst http://news.bbc.co.uk is accessible in Iran, http://www.bbc.co.uk/persian/ is not. News.bbc.co.uk was on the ONI list, whereas the Persian-language BCC World Service page was 'discovered' by the Issuecrawler. Had the 'by site' crawl setting been employed, as opposed to the 'by page' setting, only http://www.bbc.co.uk would have been retained, and that page is not blocked in Iran.

In the censored Website discovery procedure with the Issuecrawler, 30 sites were newly discovered to be blocked in Iran (see figure one), after an ONI testing procedure using proxy servers. Here one fetches pages through machines located in the country.

Initially the Internet censorship work with the Issuecrawler concerned itself with providing a practical map for Internet users in a censored country. In 1991, prior to the emergence of the Web, a phrase circulated that attributed to the Internet particular built-in properties: "The Internet treats censorship as a malfunction and routes around it."[2] In a commentary on such an idea, the mapping project sought to examine whether one could surf around, or 'circumnavigate,' censorship. A set of women's rights URLs, provided by the ONI, were inserted into the Issuecrawler, a map was generated, and the URLs were checked for blocking. To demonstrate the concept of censorship circumnavigation, the map shows that particular women's rights sites are blocked, and thus also those sites' links to other women's rights sites, but by another route, one may reach the links.

[2] The phrase is attributed to John Gilmore.

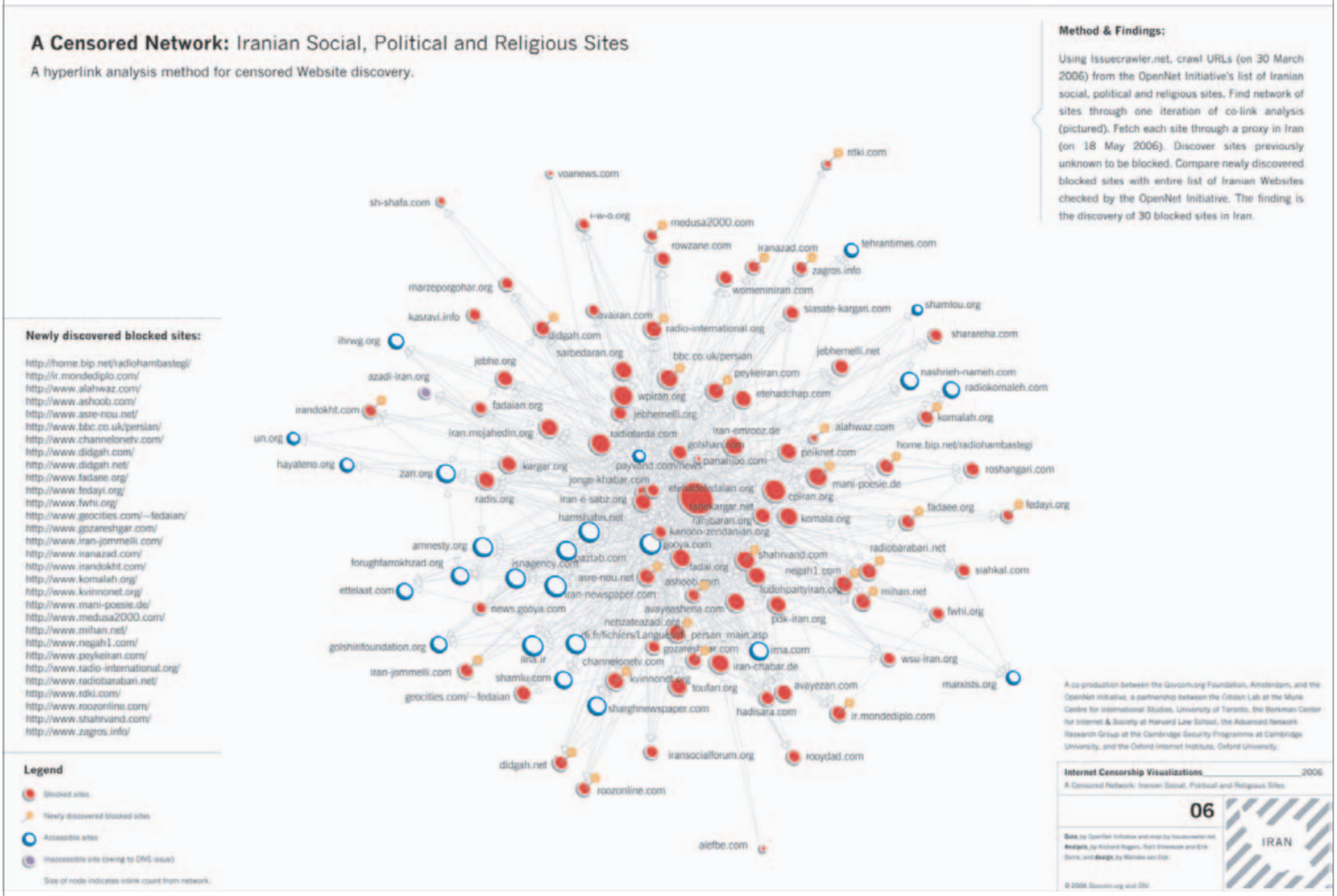


Figure Seven: Transparent proxy servers indicate that http://www.bbc.co.uk is accessible in Iran, and that http://www.bbc.co.uk/persian is forbidden.
Output from http://tools.issuecrawler.net/beta/proxies.

site fetched	response code for request	country of proxy	type of proxy	proxy	time of access
http://www.bbc.co.uk	HTTP/1.0 200 OK	Iran	transparent	85.185.236.229:3128	08/16/06 16:14:06

---

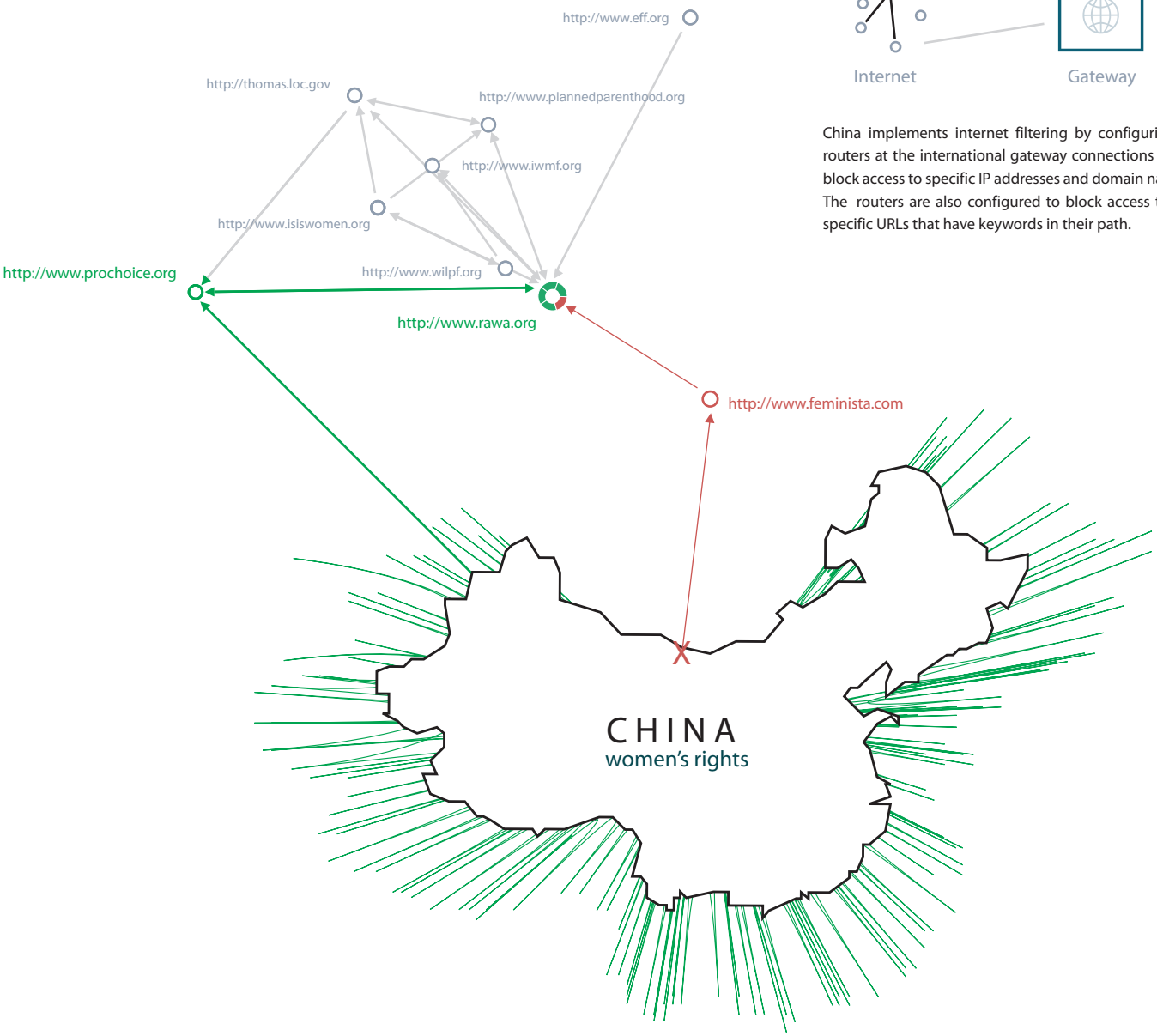
site fetched	response code for request	country of proxy	type of proxy	proxy	time of access
http://www.bbc.co.uk/persian	HTTP/1.0 403 Forbidden	Iran	transparent	85.185.236.229:3128	08/17/06 16:21:00

Figure Eight: A Censored Network - Iranian Social, Political and Religious Site: A hyperlink analysis method for censored Website discovery, Govcom.org and the OpenNet Initiative, 2006.

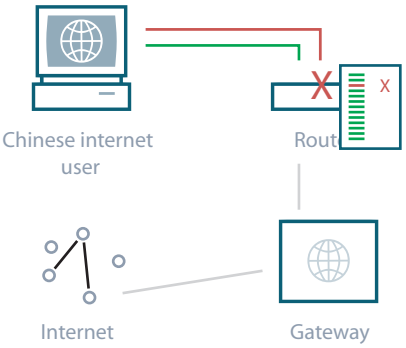
Data collection technique:
Dynamic sampling of related URLs from master lists (by category).

Each list of URLs per category is entered into a Web crawler, which locates additional sites in the same or in related categories in two different methods. 1) Locate the 'web sphere' surrounding the sites through a snowball method, capturing outlinks from seed URLs. 2) Locate a relatively denser network around the sites through co-link analysis of outlinks from seed URLs.

Results of dynamic sampling are loaded into the circumnavigation map.



Internet Censorship in China



China implements internet filtering by configuring routers at the international gateway connections to block access to specific IP addresses and domain names. The routers are also configured to block access to specific URLs that have keywords in their path.

Figure Nine: The Internet Treats Censorship as a Malfunction and Routes around it?, Govcom.org and Andermedia.nl, 2005.

Case Study: Ranking Actors (and their policy impact) with the Issuecrawler (and allied tools)

The media reform movement in the United States, which meets annually at the National Conference for Media Reform hosted by Free Press, concerns itself with such issues as media concentration and fake news (see Figure Ten). Among the sub-networks of actors convening at the events are 'media justice' groups, whose 'justice' language is part of a larger movement associated most readily with the World Social Forum and it national affiliates. Groups in the media justice area include the Funding Exchange (http://www.fex.org), Media Alliance (http://www.media-alliance.org), Media Tank (http://www.mediatank.org), United Church of Christ (http://www.ucc.org) and Youth Media Council (http://www.youthmediacouncil.org), each of which has received grants from the Ford Foundation's electronic media policy portfolio, and thus are viewed by the major funder as actors with a potential impact on media policy and reform in the United States. The question concerns how to measure policy impact. Here impact is thought of in a scientometric sense, i.e., citations by leading policy actors, whether positive or negative. Instead of pre-defining the policy actors, the heuristic concerns having the media justice actors lead the researcher to them, using the Web (and hyperlink analysis). The five URLs (above) are inserted into the Issuecrawler, and one iteration of method is chosen so as to find the immediate neighborhood around the activists. The 'by page' setting is chosen, in order to allow the advocates and activists (via the Issuecrawler) to point to specific pages on the Web, where the major issue (terms) are eventually sought in a further procedure, ultimately, to measure actor policy impact on specific issues, described below. The network (scheduled to run monthly, January through August, 2006) is initially small, comprising 20 actors (in January), 19 of which are non-governmental organizations (NGOs), including a few independent media organizations such as Indymedia, Democracy Now! and Prometheus Radio. (One iteration of method is useful in any networks containing indymedia, for indymedia links heavily to its own network of sites and thus often takes over the network, as has been the case with the September 11th Truth Movement crawl set, run by a researcher at the Annenberg School, University of Pennsylvania.)

The deep pages in the network are consulted for major issues, initially on the clickable map and subsequently on the accompanying xml file, which the Issuecrawler's visualization modules consult when rendering maps.[3] The xml file is checked because pages associated with the sites are ranked by inlink count, and the page of a site receiving the most inlinks becomes the clickable page on the map. Other deep pages may be in the network, however. Once identified, deep page (content) analysis is performed, where in the media justice network it was found that "big cable," "community media empowerment," and "broadband as public service" were the major issues in early 2006 (see Figure Eleven.)

[3] Deep pages on the clickable Media Justice – Core Actors map (and the issue language found), February 2006: http://www.media-alliance.org/article.php?story=20050331213315547 (big cable) and http://www.ucc.org/ocinc/mep/orgman.htm (community media empowerment).

Figure Ten: Action Issues List at Free Press, freepress.net, 16 August 2006.

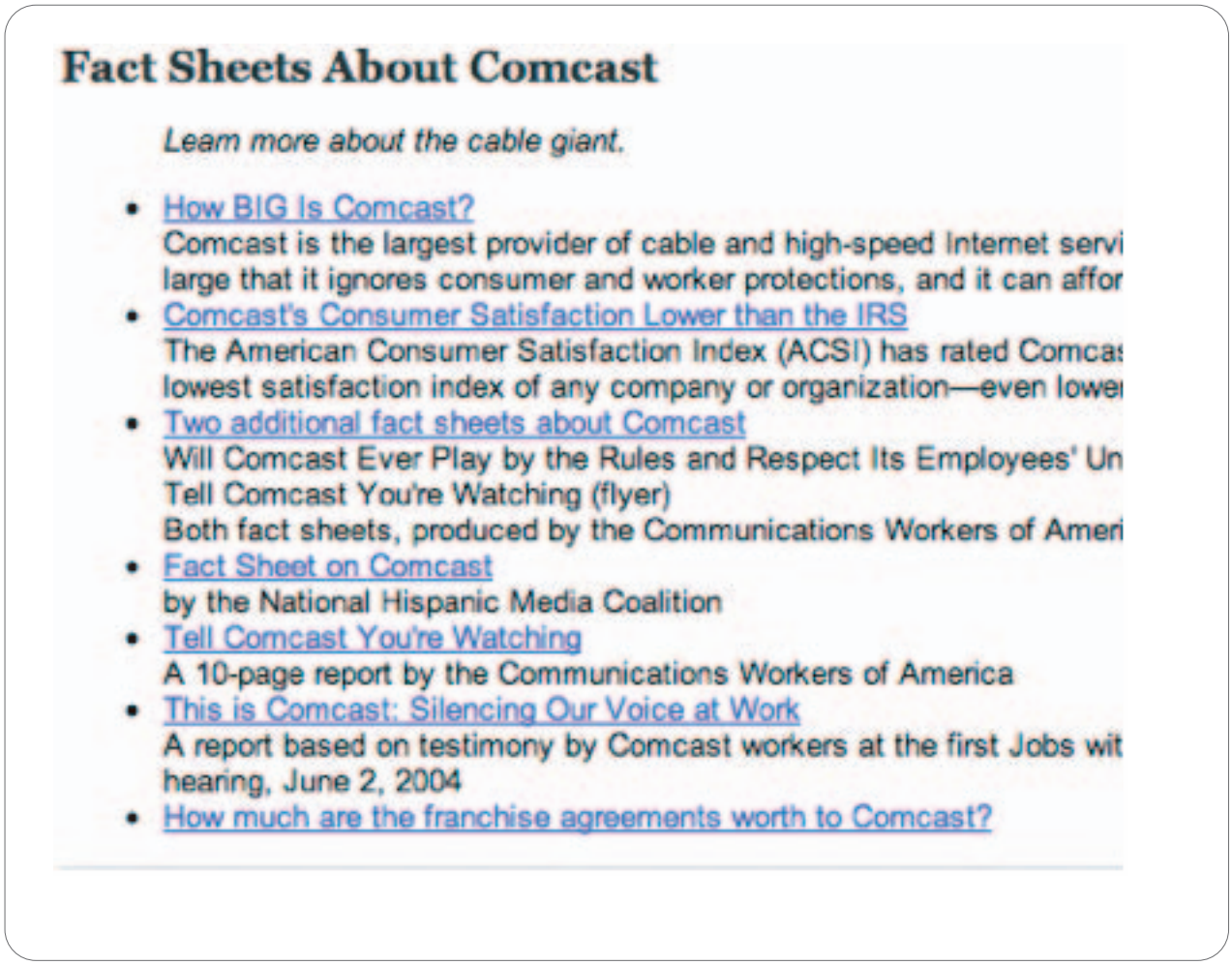


Figure Eleven: Deep Page Issue by Media-alliance.org, Feburary, 2006, formulated as 'big cable.'

For the policy impact analysis, the initial question concerns whether the network has identified a policy target. In the case of the media justice network, the target is clearly in view (see Figure Twelve). The Federal Communications Commission (fcc.gov) is the only governmental actor to receive links from the network. Thus, for the policy impact measure, that actor is queried for mentions of the media justice actors, initially, and subsequently for the media justice actors together with the issues. Finally, the issues alone may be queried to compare relative mentions of media justice actors and issues (big cable, broadband, etc.) against the universe of mentions.

In order to query the policy target (fcc.gov), one may use a leading search engine, or a tool built on top of it, at http://tools.issuecrawler.net/beta/ (see Figures Thirteen and Fourteen). Here the Google Scraper is employed, where one may perform multiple queries simultaneously and receive the citation (or mentions) data, comma separated, ready for importing ('paste special') into a spreadsheet or other software. Fcc.gov is queried for each formal organizational name in the media justice network (as governmental actors tend to use formal names), and a simple ranked actor policy impact list is created (see Table One).

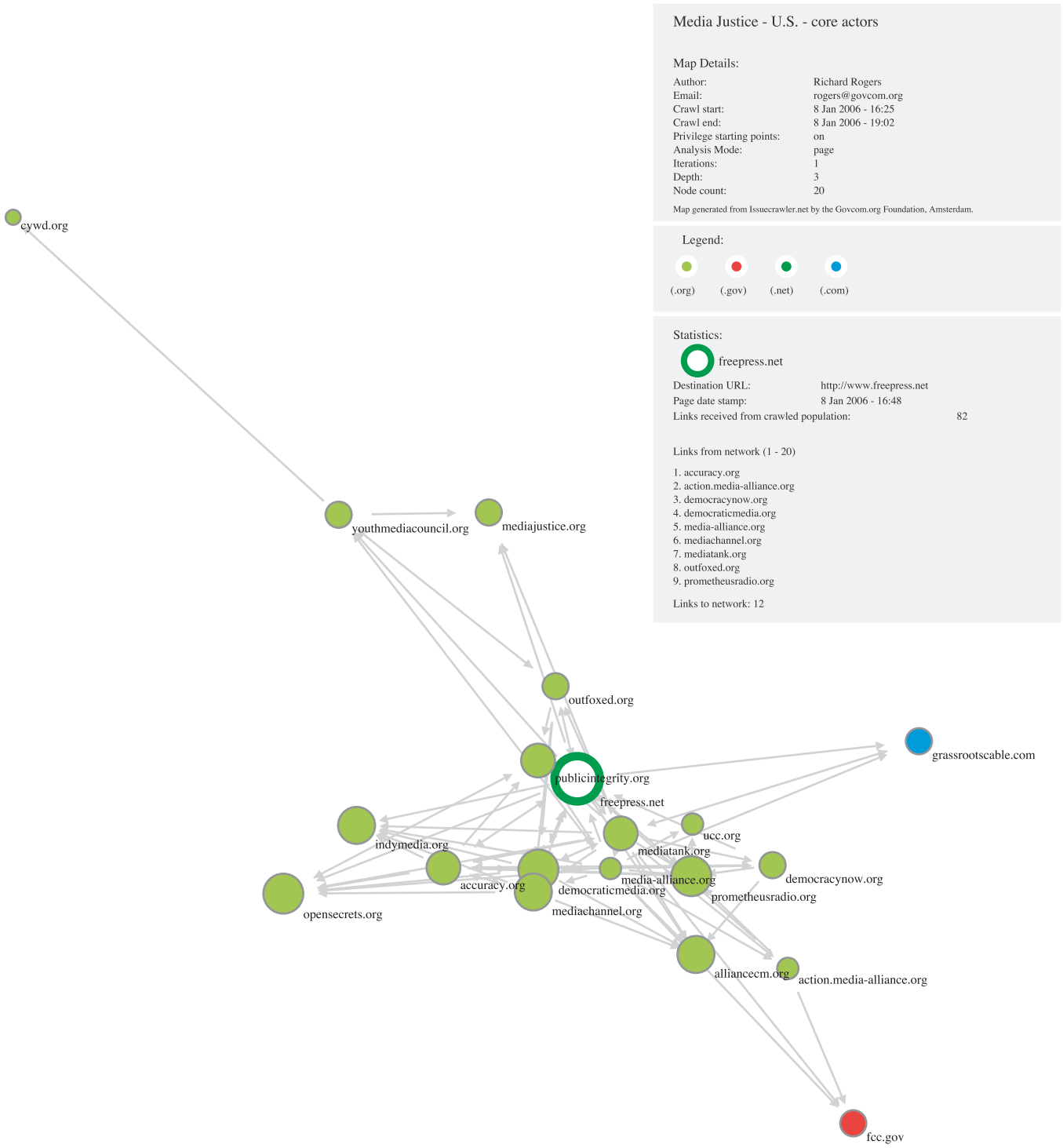


Figure Twelve: Policy Target (Fcc.gov) Identified by Issue Network. Cropped Issuecrawler Map, Media Justice Network, January 2006.

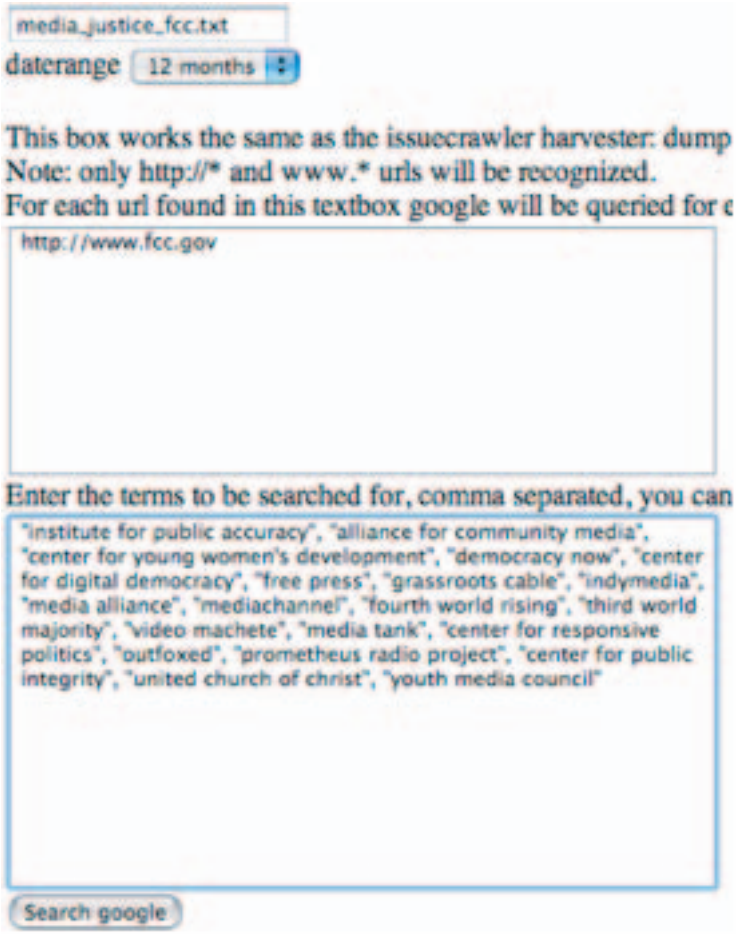
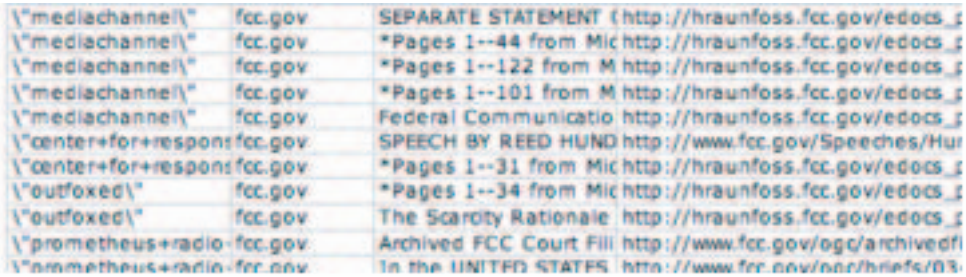


Figure Thirteen: Query fcc.gov for Mentions of Actors in the Media Justice Network, http://tools.issuecrawler.net/scrapeGoogle/, March 2006.



\"mediachannel\"	fcc.gov	SEPARATE STATEMENT (	http://hraunfoss.fcc.gov/edocs_
\"mediachannel\"	fcc.gov	*Pages 1--44 from Mic	http://hraunfoss.fcc.gov/edocs_
\"mediachannel\"	fcc.gov	*Pages 1--122 from M	http://hraunfoss.fcc.gov/edocs_
\"mediachannel\"	fcc.gov	*Pages 1--101 from M	http://hraunfoss.fcc.gov/edocs_
\"mediachannel\"	fcc.gov	Federal Communicatio	http://hraunfoss.fcc.gov/edocs_
\"center+for+respons	fcc.gov	SPEECH BY REED HUND	http://www.fcc.gov/Speeches/Hu
\"center+for+respons	fcc.gov	*Pages 1--31 from Mic	http://hraunfoss.fcc.gov/edocs_
\"outfoxed\"	fcc.gov	*Pages 1--34 from Mic	http://hraunfoss.fcc.gov/edocs_
\"outfoxed\"	fcc.gov	The Scarcity Rationale	http://hraunfoss.fcc.gov/edocs_
\"prometheus+radio	fcc.gov	Archived FCC Court Fili	http://www.fcc.gov/ogc/archivedfi
\"prometheus+radio	fcc.gov	In the UNITED STATES	http://www.fcc.gov/ogc/briefs/03

Figure Fourteen: Cropped Output of Query of Fcc.gov for Media Justice Actor Mentions, http://tools.issuecrawler.net/scrapeGoogle/, March 2006.

Table One: Media Justice Actor Rankings by Frequency of Mentions by the FCC, March 2006.

1.	Center for Digital Democracy	(78)
2.	Alliance for Community Media	(61)
3.	United Church of Christ	(60)
4.	Prometheus Radio	(56)
5.	Media Alliance	(23)
6.	Free Press	(21)
7.	Democracy Now	(12)
8.	Media Channel	(7)
9.	Grassroots Cable	(5)
10.	Center for Responsive Politics	(2)
10.	Outfoxed	(2)
12.	Indymedia	(1)

Note: Other media justice actors in network not mentioned by the FCC.

The overall policy impact measure is supplemented by the policy issue impact measure. One queries the policy target for actor associations with issues. Here the actors and "broadband" are queried at the FCC as well as those and "big cable." To monitor collective network actor impact on policy target, the simple percentage measure is displayed. For "big cable," 8/19 actors are mentioned, for "broadband," 14/19. A further quantitative measure has broadband with the actors mentioned with much greater frequency at the FCC than "big cable" or "community media."

In analyzing network growth over time, the crucial measure for an individual actor is change in inlink count. For specific networks, monitored for example by funding agencies, significant change in actor inlink count is similar to a rise or fall in one's search engine position for e-commerce and other actors. The drama of search engine placement drop, or precipitous rise, in the case of the *New York Times on the Web* (from February to April, 2005) may be depicted in graph form (see Figure Fifteen.) "WWW" was queried in leading search engines (each with distinct logics) over a period of time, and the results were triangulated and plotted to graph. A rendition of the overall hyperlink economy, albeit just top sites, was produced. The Issuecrawler, however, concerns itself with micro-hyperlink economies.

Hyperlink Economy

Triangulated results of 'WWW' query in Google, Yahoo!, MSN and Teoma,
28 February to 26 April 2005, with profiles of top actors from
the categories: open source organizations, computer companies,
search engine companies, non-computer companies, news companies,
non-governmental organizations and governmental bodies.

Govcom.org Issue Mappings _____ 2005
Hyperlink Economy

WSIS edition **05**

Thanks to the Social Science Research Council's Information Technology and International Cooperation Program.

Analysis_Dragana Antic

Product of the thematic project," Web Epistemology, Information Politics and Augmented Reality,"
Piet Zwart Institute, Rotterdam.

SSRC

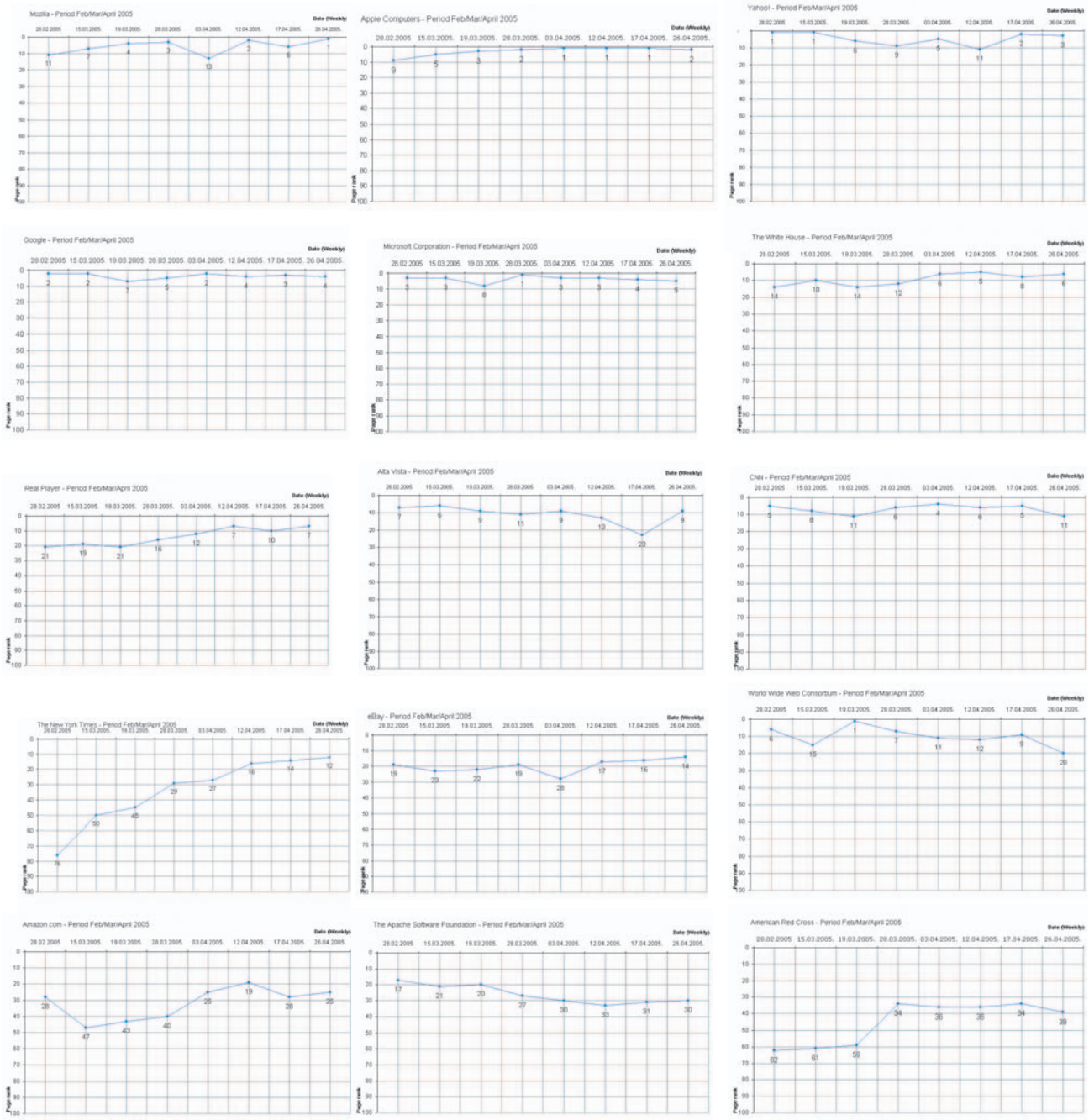


Figure Fifteen: Hyperlink Economy, 2005. Analysis by Dragana Antic, Piet Zwart Academy, Rotterdam, with Richard Rogers. Design by Marieke van Dijk.

Changes in an actor's position in the network over time is performed with an Issuecrawler allied tool, http://tools.issuecrawler.net/beta/comparenetworksovertime/. For example, in the media justice network, to depict actor ranking moves over time (from January to August, 2006), one inputs the xml files of each result of the scheduled crawls. The 'baseline' option shows changes against one time period, whereas the 'previous network' setting outputs month-to-month changes. Modifications in site design, e.g., an upgrade from an html site to a content management system, as has been the current norm, may account for sudden changes to actors' inlink counts, however much it is important to note that decisions are taken when modifying sites actually to remove links. Link removal ends a relationship on public display (the classic link list), though it may be renewed on day-to-day postings, blog-style, should an actor's campaign, event, tool, press release or other form of information release once again merit a link. Thus the Issuecrawler downplays the overall importance of site design modification, and continues to focus on an actor's information actions and recognition of them by other actors. For example, in the media justice network, media-alliance.org received its most inlinks, month-to-month, for separate deep pages. It is the only actor which can make such a claim, though over time grassrootcable.com saw its linkages move from homepage to deep page, grassrootcable.com/fcc.html, where one may file a comment to the FCC about national cable ownership limits.

An original design for an Issuecrawler actor profile output takes into account types of inlinks an actor has received from the network (see Figure Two). The quantity as well as type of inlinks received, when analysed over time, show changes in standing. In the case of the media justice network, no actor receives links from the FCC at any time, however. An outlink stripper that scrapes a single site for every outlink, and characterizes the organization on the basis of its outlinks proves useful in this respect (http://tools.issuecrawler.net/beta/linkRipper/). Here the point is to enquire into whether an actor, as the FCC, ever links to .org's, and to which .org's, doing which issues. Should the FCC link to .org's engaged in 'big cable' and/or 'broadband,' without, as is the case, recognition for media justice actors, one has another relative policy impact indicator.

Case Study: The Extended Event Network – Map Interpretation

Why hold an event? In the NGO-funder's community, where, apart from science, much Issuecrawler user activity is situated, often funders will suggest to NGOs that two types of groups should 'network,' for example, the open source community and NGOs, convened on the Island of Vis, Croatia, by the Tactical Technology Collective in the Summer of 2002, with support from the Open Society Institute, Budapest. Funders may subsequently look at Issuecrawler maps to endeavor to view, through linking behavior, the extent of the networking, before and after the event.

In the case of the hackers' event, What the Hack, held in a rural setting in the Netherlands in July 2005, of interest to the researchers was the linking behavior hackers and other linked-to groups would exhibit, and how it could be interpreted. The procedure for mapping an event network often entails hanging a sheet at the reception table, with the words written at the top, "Who's Here? URLs please" (see Figure Sixteen). After a suitable period of URL collection time, the URLs are typed out, and entered into the Issuecrawler, with one iteration of method and privileged starting points, so as to keep as many starting points on the map, and also to display those parties recognized by the attendees by links (which the researchers often dub, Who should be here (as well)?) At the What the Hack festival, the sheet-hanging technique was successful, and eventually an extended event network map was made. Hacker starting points were entered. (Appropriately, gnu.org, founded by Richard Stallman, president of the Free Software Foundation, which issued the General Public License, is the top node.) Two groupings emerged on the map, the hackers as well as an information policy actor set, including the Electronic Frontier Foundation (eff.org), the Electronic Privacy Information Center (epic.org) as well as European Digital Rights (edri.org).

Straightforward interpretations of issuecrawler maps of the cluster variety may begin by recognizing whether actor types cluster. The researchers' interpretation of the What the Hack map followed from noting two significant clusters -- the information policy grouping -- including the actors mentioned above as well as Privacy International and the Global Internet Liberty Campaign (gilc.org) -- and an open source grouping -- including dynebolic.org, freaknet.org and a set of dyne.org pages, many related to the media activist open source bootable operating system and software CD. It is evident on the map that the open source cluster is distant from the information policy actors (see Figure Seventeen). Intriguingly, the event site, whatthehack.org, is closer to the information policy grouping than to the open source cluster, suggesting that the event itself is considered more important (by the information policy people) than the open source community working on a daily basis. The map appears to show that events such as What the Hack (as opposed to everyday practice) are an object of information policy interest.

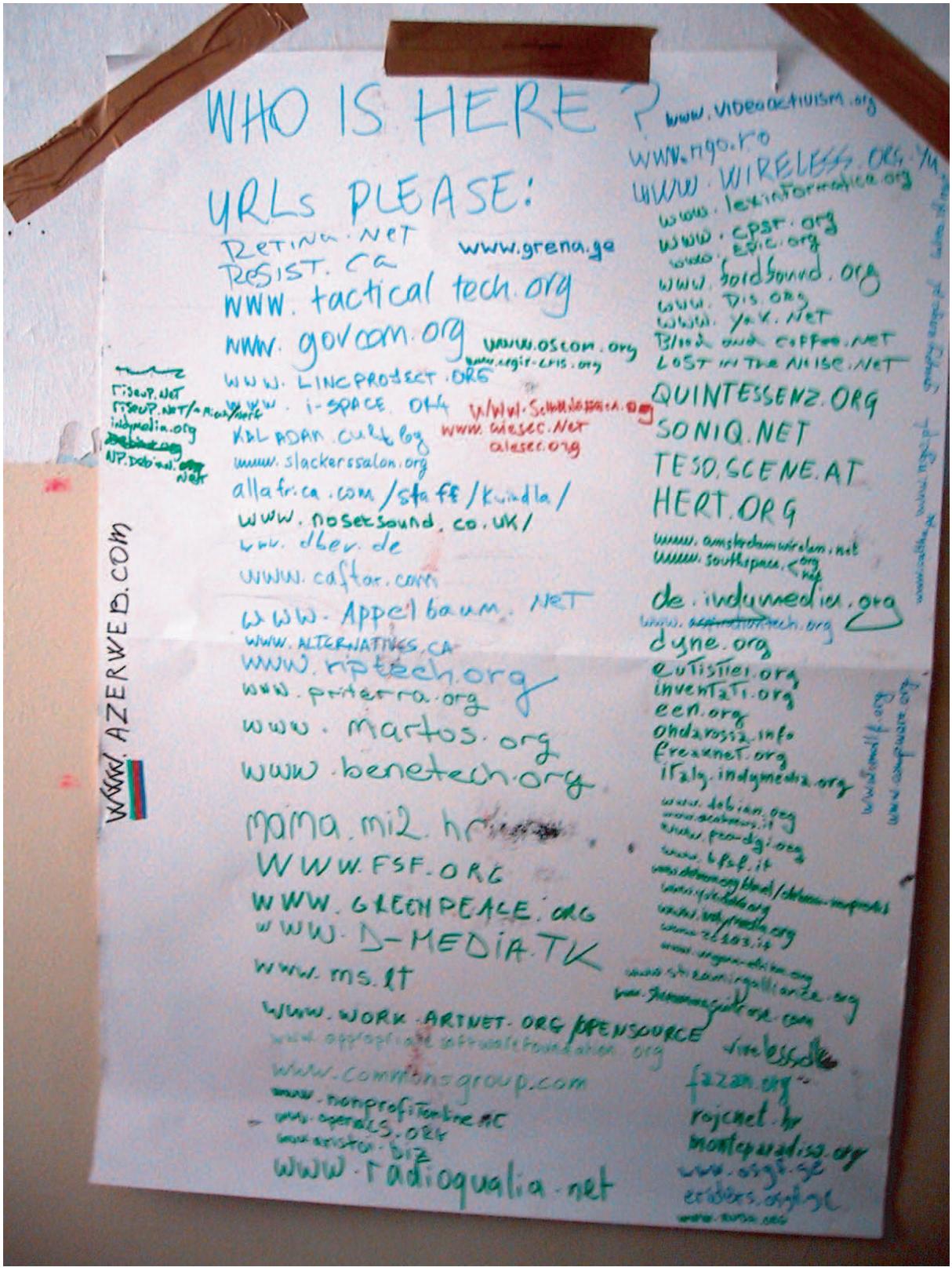


Figure Sixteen: URL Collection Technique at an Event. Summer Source camp, Island of Vis, Croatia, 2002.

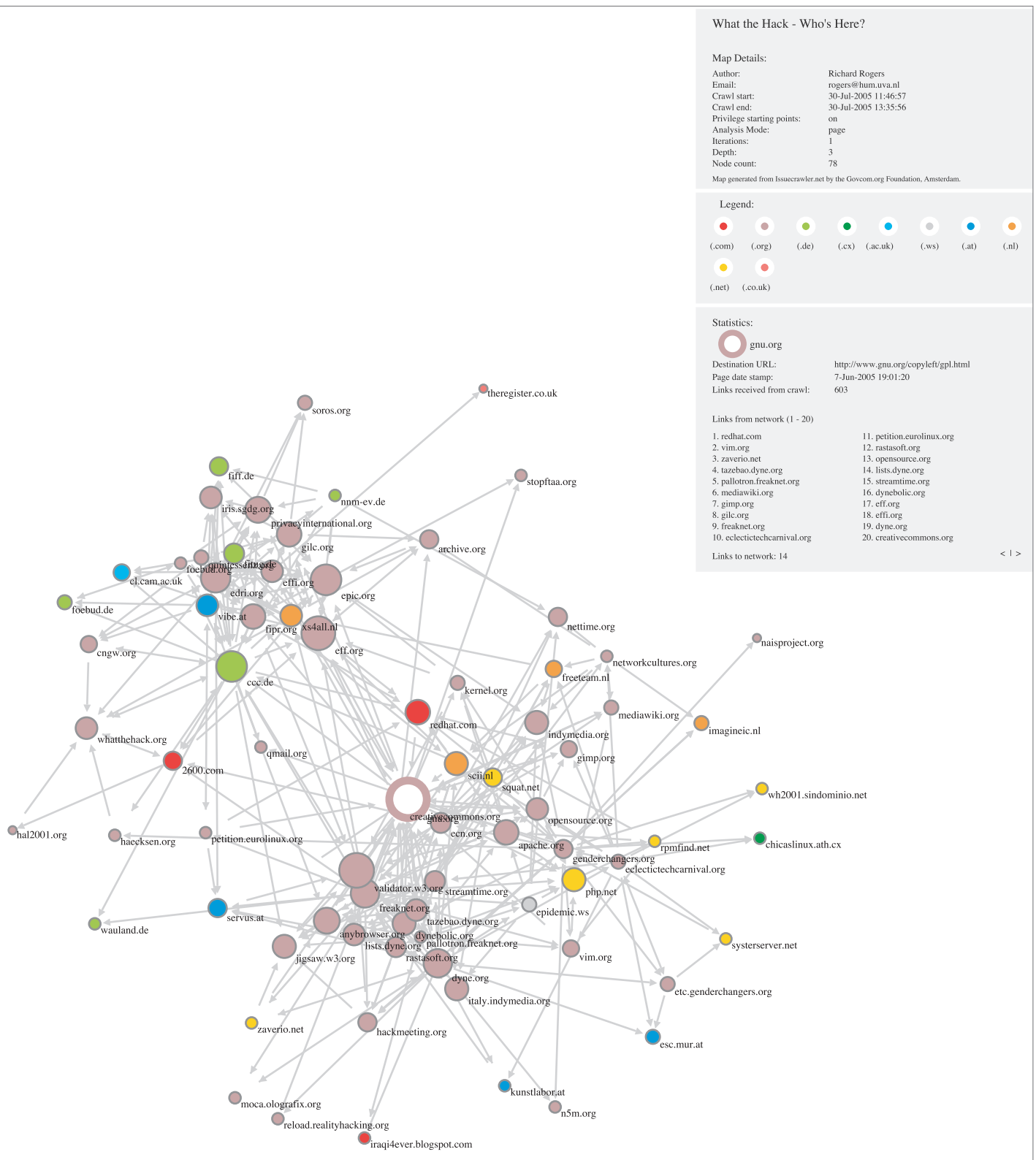


Figure Seventeen: Information Policy Actors Cluster Around Hacker Event Website. What the Hack Extended Event Network, Govcom.org with Anderemedia.nl and Sonologic.nl, 2005.

Conclusions

Issuecrawler visualizations have evolved with conceptualizations of Web space -- from hyperspace and cyberspace over public sphere and debate space to network and locative media. In each case the visualizations sought to engage with specific notions of Web space. In the hyperspace period the Netlocator tethered sites by showing inlinks. The Issuecrawler broke with the alleged open-ended-ness of cyberspace by showing how hyperlinks demarcate associational space. It also engaged with public sphere theory (Web as debate space) by unflattening the virtual roundtable, showing over-sized nodes and entangling links. The cluster map module organized actors into a particular kind of network, the issue network, where, with the Issuegeographer, one is able to map the distance between where an issue is happening (e.g., on the ground), and where an issue is currently based (e.g., in a summit network).

Recent concrete research projects with the Issuecrawler additionally engage with the current locative media period, where the Web, with tools, may be made to show information politics in very specific geographical settings -- a hacker camp in a field in the Netherlands, a U.S. governmental agency in Washington, D.C., the national censorship policy of Iran. The Web tells us that hackers interest information policy circles by holding an event, that there is a specificity to the media justice issues and actors recognized by the U.S. Federal Communications Commission, and that Iran blocks the Persian-language page of the British Broadcasting Corporation.

Bibliography

[BAR96] Barbrook R., Cameron, A., "The Californian Ideology," Science as Culture, vol. 6, nr. 1, p. 44-72, 1996.

[BER99] Berners-Lee T., *Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor*. London: Orion, 1999.

[BUS45] Bush V., "As We May Think," *Atlantic Monthly*, p. 101-108, July, 1945.

[DAI02] *The Daily Telegraph*, 30 November 2002.

[DEA02] Dean, J., *Publicity's Secret: How Technoculture Capitalizes on Democracy*, Ithaca, NY: Cornell University Press, 2002.

[DOD01] Dodge M., Kitchin, R., *Mapping Cyberspace*, London: Routledge, 2001.

[GOV99] Govcom.org, *The Rogue and the Rogued: Amongst the Web Tacticians*, Jan van Eyck Design and Media Research Fellowship, Maastricht, 1999 [video].

[GOV05] Govcom.org, *The Places of Issues: The Issuecrawler Back-end Movie*, prepared for "Making Things Public: Atmospheres of Democracy," curated by B. Latour and P. Weibel, ZKM, Karlsruhe, Germany, 2005 [movie].

[HOB03] Hobbs R., *Mark Lombardi: Global Networks*, New York: Independent Curators International, 2003.

[KEH99] Kehoe C., Pitkow J., Sutton K., Aggarwal G., Rogers J., *GVU's Tenth World Wide Web User Survey*, Graphics Visualization and Usability Center, College of Computing, Georgia Institute of Technology, Atlanta, Georgia, 1999.

[MAR99] Marres N., Rogers R., "To Trace or to Rub: Screening the Web Navigation Debate," *Mediamatic*, vol. 9/10, p. 117-120, 1999.

[MAR00] Marres N., "Somewhere you've got to draw the line: De politiek van selectie op het Web", MSc thesis, University of Amsterdam, August 2000.

[MAR05] Marres N., *No Issue, No Public: Democratic Deficits after the Displacement of Politics*, PhD Dissertation, University of Amsterdam, 2005.

[MAR06] Marres N., "Net-Work Is Format Work: Issue Networks and the Sites of Civil Society Politics", In: Dean, J., Asherson, J., Lovink, G. (eds.), *Reformatting Politics: Networked Communications and Global Civil Society*, 2006.

[MIL01] Miller D., Slater D., *The Internet: An Ethnographic Approach*, Oxford: Berg, 2001.

[NEL99] Nelson T., "Xanalogical Structure, Needed now more than ever: Parallel Documents, Deep Links to Content, Deep Versioning, and Deep Re-use", ACM Computing Surveys , vol. 31, issue 4es , December 1999.

[ONI06] OpenNet Initiative, "Internet Filtering in Vietnam in 2005-2006: A Country Study", University of Toronto, Harvard University, Cambridge University, 2006.

[PAI05] Pain J., "Internet-censor world championship," In: Devilette S. (ed.), *Handbook for Bloggers and Cyberdissidents*, Paris, France: Reporters without Borders, 2005.

[PAR05] Park H., Thewall, M., "The Network Approach to Web Hyperlink Research and its Utility for Science Communication", In: Hine C. (ed). *Virtual Methods: Issues in Social Research on the Internet.* Oxford: Berg, p. 171-181, 2005.

[RHE94] Rheingold H., *The Millennium Whole Earth Catalog*, San Francisco: Harper, 1994.

[ROG96] Rogers R., "The Future of Science & Technology Studies on the Web", *EASST Review*, vol. 15, nr. 2, p. 25-27, June 1996.

[ROG00] Rogers R. (ed.), *Preferred Placement: Knowledge Politics on the Web*, Maastricht: Jan van Eyck Editions, 2000.

[ROG04a] Rogers R., "The Departure of Science from the RFID Issue Space: Mapping the State of an Emerging Sociotechnological Controversy with the Web", paper presented at the EASST Conference, Paris, 28 August 2004.

[ROG06] Rogers, R., *Information Politics on the Web*, Cambridge, MA: MIT Press, 2006.

Acknowledgments

The author would like to thank Noortje Marres for critical commentary. The Iranian censorship case study benefited from collaboration with Nart Villeneuve of the OpenNet Initiative, University of Toronto. The Media Justice case was supported by the Electronic Media Policy portfolio of the Ford Foundation, New York, facilitated by Becky Lentz. Thanks to Anat Ben-David, Erik Borra, Marieke van Dijk and Koen Martens for the map-making at the hacker event in the Netherlands. Appreciation is extended to Auke Touwslager for designing the Chinese censorship circumnavigation graphic and to Dragana Antic for conducting the hyperlink economy research.